A numerical database for ultrasonic defect characterisation using array data: robustness and accuracy

Authors: Robert Malkin*, Mihai Caleap and Bruce W Drinkwater

* Corresponding author

Email addresses: r.e.malkin@gmail.com, mihai.caleap@bristol.ac.uk, b.drinkwater@bristol.ac.uk

Address: Department of Mechanical Engineering, University of Bristol, Bristol, UK, BS8 1TR

Abstract

Elastodynamic scattering matrices are known to contain geometrical information about a given scatterer, such as its size and shape. Here, the extent to which this scattered information can be retrieved using an ultrasonic array and used to characterise defects for Non-Destructive Evaluation is explored. Experimentally measured defect scattering matrices are compared to a database of possible scatterers and the nearest neighbour used to characterise the defect's geometry in terms of crack length and orientation. As an example, a database of scattering matrices for small (lengths 0.2-2.0mm) cracks at a range of frequencies (2-20MHz) is formed. The short range similarity (i.e. that between close neighbours) and the long range similarity (i.e. uniqueness) are used to understand the uncertainties inherent in this approach. In addition, the effect of spatially coherent noise, such as grain scattering in a polycrystalline metal, on the scattered information content is quantified. It is shown that as the noise level or frequency increases, so the information retrievable from a given crack is reduced, setting bounds on the accuracy of characterisation possible from a given ultrasonic dataset.

1 Introduction

In order to assess the safety of a structure it is important to be able to detect, locate and characterise various types of defects. Ultrasonic array-based Non-Destructive Testing (NDT)

techniques have allowed for this to become a regular aspect of structural inspections. Of particular interest is the ability of an inspection to fully characterise the nature of the defect and to decide if it is a critical defect which may lead to failure, or if the defect is sub-critical and may safely be ignored or monitored.

Using ultrasonic array imaging methods such as the: total focusing method [1], synthetic aperture focusing technique [2], [3], inverse field wave extrapolation [4] or wavenumber algorithms [5], a range of defects can be detected and accurately located. If the defects are large, i.e. greater than a few wavelengths, these methods also offer the possibility to determine the defect shape and/or orientation of the defect directly from the image [6]. Sizing of large defects is also possible using mathematical models of scattered signal data [7], corner trap the tip diffracted echoes [8] and superimposed echo modelling [9], [10]. If the defect is small, it has been shown that characterisation may be possible by utilising information contained within the defect scattering matrix, which describes the angular scattering behaviour of a given scatterer [11].

The scattering matrix, or S-matrix, describes the amplitude and phase of the scattered field of a defect in the far field, and has been shown to encode the far-field information arising from all wave-scatterer interactions [12]. Let $\boldsymbol{r}$ be the position vector of a point in the x-z plane which in polar coordinates is given by, $\boldsymbol{r} = (r, \theta)$; here $r = |\boldsymbol{r}|$ and θ is measured from the positive z-axis. For 2-D problems, the far-field scattering amplitude is defined by, [13].

$$u^{\mathrm{sc}}(\boldsymbol{r}) \simeq \sqrt{\frac{2}{\pi}} e^{-i\pi/4} \frac{e^{ikr}}{\sqrt{kr}} \; S(\theta) \quad \text{as} \quad r \to \infty, \tag{1}$$

where $i = \sqrt{-1}$, $k = \omega/c_L$ and c_L is the longitudinal wave velocity. For a given angular frequency, ω, $S(\theta, \omega)$ gives the field scattered in the direction θ. If the incident field is a plane wave propagating in the direction θ_{in}, we write $S(\theta, \theta_{in}, \omega)$ – this angular shape function is what we call in our paper the S-matrix. Note that here we approximate the far-field array element output in the vicinity of a small defect as a plane wave.

The S-matrix, or more strictly, portions of it, may be extracted from array data using, for example, back-propagation imaging as shown in [14]. As S-matrices are extracted from a known point within an array image all the propagation paths/distances are known, therefore propagation, directivity and frequency dependent damping effects may be accounted for. Zhang et al. [6] demonstrated that if the extracted S-matrix contains a specular reflection, accurate sizing of a 1-D crack (for crack lengths $\approx 0.5\text{-}1.5\lambda$) is possible. A recent study by the authors used a database of pre-calculated S-matrices against which an experimentally extracted S-matrix was compared [11]. Using this approach, the length and orientation of cracks could be accurately measured, even if the specular reflections were not captured. This work showed that even relatively small regions of the S-matrix (i.e. in terms of range on incidence and scattered angles) potentially contain sufficient information for defect characterisation. It should be noted that realistic defects will likely feature surface roughness, for which a database of pre-calculated S-matrices would be unmanageably large. Recent work has demonstrated that defects roughness can be thought of as adding noise to the defect classification process, i.e. classification is unaffected by low levels of surface roughness and, as the roughness increases, so does classification uncertainty [15]. We build on this work by exploring the information contained within an S-matrix and the measurement confidence of the characterisation result. In particular, we examine the effect of spatially coherent scattering noise on S-matrix information content and hence on defect characterisation accuracy.

The Structural SIMilarity (SSIM) index was originally developed as a method of objective evaluation of the similarity between two optical images, a reference and modified/degraded image. SSIM compares two images (X, Y) based on three parameters; mean, $m(X, Y)$, variance, $v(X, Y)$ and cross-correlation, $q(X, Y)$. The SSIM between the two images X and Y is given by

$$\begin{aligned} SSIM(X,Y) &= m(X,Y)^{\beta} \cdot v(X,Y)^{\gamma} \cdot q(X,Y)^{\delta} \\ &= \left(\frac{2\mu_X\mu_Y+C_1}{\mu_X^2+\mu_Y^2+C_1}\right)^{\beta} \cdot \left(\frac{2\sigma_X\sigma_Y+C_2}{\sigma_X^2+\sigma_Y^2+C_2}\right)^{\gamma} \cdot \left(\frac{\sigma_{XY}+C_3}{\sigma_X\sigma_Y+C_3}\right)^{\delta}, \end{aligned} \tag{2}$$

where, μ is the mean, σ the standard deviation, σ_{xy} the cross-correlation (inner product) and $C_{1,2,3}$ are small constants used to avoid numerical instabilities when $(\mu_X^2 + \mu_Y^2)$ and/or $(\sigma_X^2 + \sigma_Y^2)$ and/or $(\sigma_X \sigma_Y)$ are close to zero. An SSIM of 1 results from two identical images. The exponents β, γ, δ allow the contributions of each of the three major terms to be modulated, we let $\beta = \gamma = \delta = 1$ as per the original published work [16].

2 S-Matrix database

As the number of possible defects, and thus corresponding S-matrices, is infinite we restrict our analysis to that of planar cracks of negligible thickness. Not only does this restriction have the potential to increase the chance of successful characterisation, but it also represents arguably the most important type of defect found in engineering materials. Future work will address the challenge of exploring the implications of further extending the range of defect types. The S-matrix database was created using the efficient finite element method developed in [17]. The method uses an integral representation of the wave field where the scattered field from an arbitrary shaped scatterer is decomposed into the far field scattering amplitudes, allowing the S-matrix to be calculated. The S-matrix calculation is performed at a single frequency yet typically it will be compared it to an experimental S-matrix measured using a transducer array with some bandwidth about its central frequency. This is reasonable as it has been shown that calculated single frequency S-matrices correlate extremely well with experimental S-matrices measured from time-domain data. It is also worth noting that techniques are available for extracting the S-matrices as a function of frequency and work is ongoing to understand this additional dimension to the information [14]. As shown in Figure 1, our database is created from planar crack-like defects in a two-dimensional isotropic elastic material (aluminium, with properties from [18]). The $360^{\circ}_{incident}{:}\,360^{\circ}_{scattered}$ angle S-matrices were computed with 1° angle increments. An experimentally acquired S-matrix however, will likely be significantly smaller in angular range than this due to the measured inspection angles which are possible. In order to study the effect of this reduced S-matrix each S_D was truncated into

ten sub-matrix sizes ($[36^{\circ}_{inc.}:36^{\circ}_{sca.}]$, $[72^{\circ}_{inc.}:72^{\circ}_{sca.}]$, …., $[360^{\circ}_{inc.}:360^{\circ}_{sca.}]$). As we keep the incident and scattered angular ranges equal we refer to each sub-matrix size simply by its angular range, i.e. a sub-matrix size of $36°$ = $[36^{\circ}_{inc.}:36^{\circ}_{sca.}]$. A database was then constructed with crack length, a, ranging from 0.2mm to 2.0mm in 0.03mm increments, the rotation angle, α, ranging from 0° to 178° in 2° increments (the S-matrices are identical for cracks with ±180° rotation thus we only compute half the rotational range), with all studies performed from 2MHz to 20MHz in 1MHz increments. In total, per frequency, the database contains 5400 S-matrices, individually referred to as database elements, S_D. Each S-matrix had a data size of 790 Kbytes, stored as a 360×360 32 bit number matrix of the amplitude of the complex S-matrix. The database in effect acts as a 'look-up' reference against which experimentally extracted S-matrices may be compared and the nearest neighbour found. The assumption which is explored in this paper is that, as the SSIM increases, so the nearest neighbour in the database approaches the true experimental result in terms of the characteristics of the defect.

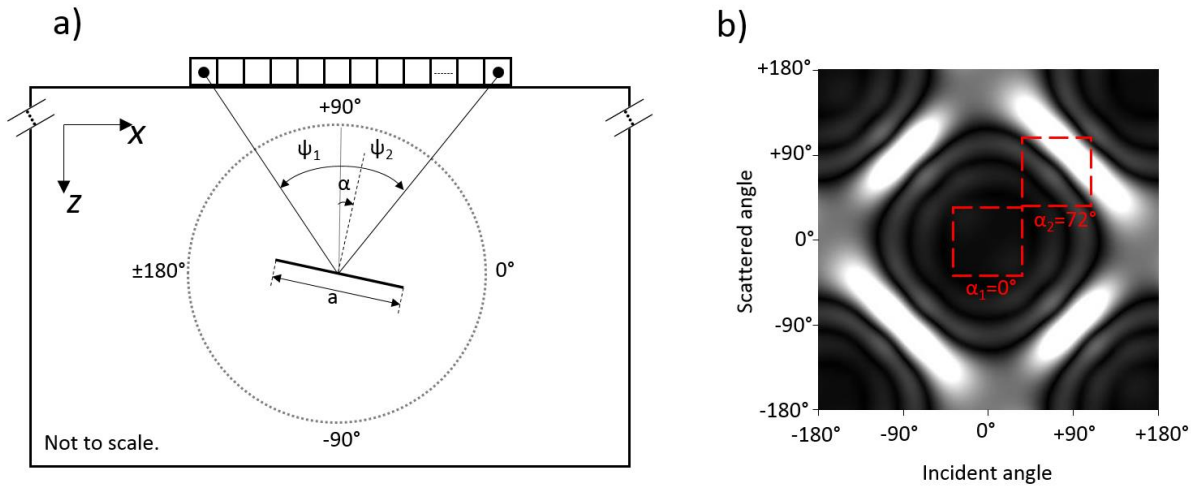

Figure 1 – Defect geometry and example S-matrix. ***a****) Geometry of crack-like defect, of length a, from which S-matrices are generated. Where ψ_1 and ψ_2 define the range of possible incident/scattered angles when using a fixed array, α is the rotation angle of the crack and x & z are the spatial coordinates.* ***b****) Example S-matrix showing the full incident:scattered angular range ±180° for a defect of a=1.00mm and α=0° at 10MHz. The red dashed boxes show 2 sub-matrices (both 72° in size) of the full S-matrix as may be acquired experimentally given limited inspection angles (ψ_1* and *ψ_2* in **a**). *The two sub-matrices represent the same defect shape with different rotations, α_1 and α_2.*

2.1 S-matrix information content

The information content of an S-matrix will depend upon the defect shape, frequency and wave type. As an example, for longitudinal waves the S-matrix of a progressively shorter crack tends to that of a point, i.e. an omnidirectional reflector which has no angular dependence. Such simple scattering behaviour can be described by a single complex number. A crack of discrete length will however have some angular dependence, as observed in Figure 2, and hence a greater number of parameters are required to describe its scattering. This increase in the number of parameters can be thought of as an increase in information content. It is of interest to know how much information is contained within an S-matrix and how this varies with defect shape and frequency. While the S-matrices we generate have almost exactly the same size when digitally stored, this does not mean that each one contains the same amount of information.

There is no accepted single objective measure of information content of a 2D image, it is however generally separated into two categories: image complexity and spatial information. Image complexity is determined by the number of unique elements within the image and the spatial information by the number of regularities found in the image [19]. An image may have high complexity but low spatial information (such as an image of randomly distributed amplitudes, e.g. white-noise) or low complexity and high spatial information (such as an image of a number of simple repeating patterns).

One measure of the complexity of a dataset is its Kolmogorov complexity, which is based on the ability to represent a dataset by the least amount of data possible [20]. As an example of the Kolmogorov complexity consider two character sequences, (a) and (b), where (a) = p,q,p,q,p,q,p,q and (b) = g,t,r,x,z,k,a,m. It is possible to represent (a) using less data with ‘p,q x 4’ whereas (b) cannot be shortened, thus (b) has a higher Kolmogorov complexity than (a). The Kolmogorov complexity however is not computable and it must be approximated using a compression algorithm [21]. A suitable algorithm being the discrete cosine transform (DCT), where the DCT expresses a dataset as a

summation of cosine functions, as featured in JPEG compression algorithms [22]. It is a good indicator of complexity as a large number of terms would be needed to describe a white-noise image but significantly less would be required for an image of repeating patterns. The image complexity metric as indicated by compression size is often used by researchers in the field of image analysis, examples shown in [23]–[25]. A suitable indicator of spatial information within an image is to measure its edge content, achieved using the Sobel filter (a 2-D spatial gradient operator which emphasises high spatial frequencies which correspond to edges) [26]. We show these metrics in Figure 2 by calculating both the image complexity (i.e. file size after compressing the S-matrix using the JPEG algorithm) and spatial information (i.e. average number of edges given upon applying the Sobel operator on the S-matrices) across a range of crack lengths and frequencies (Figure 2 and some subsequent images use linear interpolation for smoothing, interpolated figures will be indicated as such). For both we see that the information content increases with increasing crack length and frequency. From Figure 2b&c it can be seen that image complexity and spatial information are strongly correlated, as observed in [23].

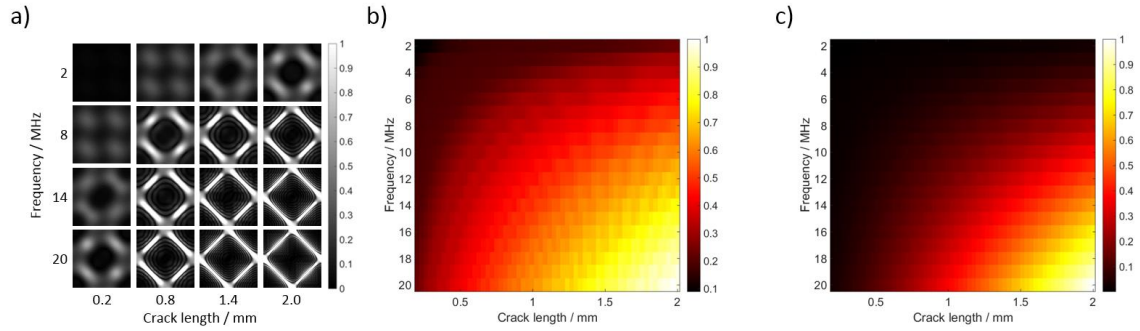

Figure 2 - S-matrix characteristics. ***a****) S-matrices (normalised amplitudes) for a range of crack lengths and frequencies (scale bars omitted for clarity, but are identical to Figure 1b).* ***b****) Image complexity: Normalised compressed file size of S-matrices showing increase in complexity with increasing crack length and frequency.* ***c****) Spatial information: Normalised average number of edges of S-matrices showing increase in spatial information with increasing length and frequency. Interpolation used for* ***b)*** *and* ***c)****.*

3 S-Matrix similarity

An important question regarding the information content of an S-matrix is how it changes with varying crack length and/or frequency. We address this by analysing the similarity of a single S-matrix when comparing it to other S-matrices. The uniqueness of an element within a group can be considered in two aspects; 1) is the element truly unique (not repeated elsewhere in the group)? and 2) is the element similar to other elements? For example: given the group [1, 2…,100] elements 7 & 8 are both unique (not repeating) and are relatively similar in magnitude, whereas 3 & 77 are unique but of dissimilar magnitude.

We can represent our S-matrix database as a 3D array with crack length, rotation and frequency as the dimensions. To compare S-matrices within this array it is useful to define a 'distance', L, between neighbouring S-matrices. Here we suggest that one single increment step in either of our crack parameters (length, rotation and frequency) is equal to a distance of 1. It is worth noting that the distance between two specific S_D is a function of the database increment size. However, this is a common approach in database searching [27].

3.1 Local similarity

Similarity between an experimental S-matrix and the database can potentially be utilised to characterise a defect, as shown in [11]. At a given frequency a single S-matrix within the database will have between 3 and 8 neighbouring S-matrices (depending on its position within the database array). The neighbouring elements of a given S-matrix will represent a crack with length and/or rotation difference of one increment (i.e. 0.03 mm or 2°). Figure 3 shows the local similarity (i.e. the SSIM averaged across neighbours of distance equal to unity) between S-matrices within the database for a range of frequencies and sub-matrix sizes.

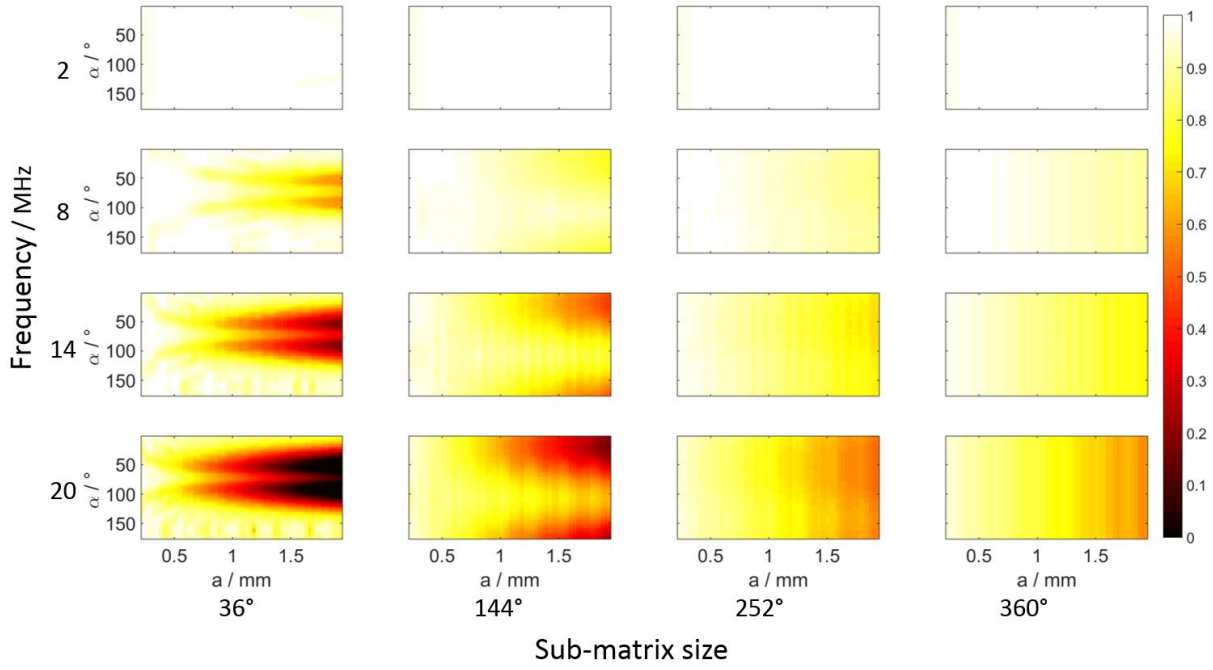

Figure 3 – Average local nearest neighbour SSIM score.

Figure 3 suggests that local similarity decreases with increasing crack length (observed at all sub-matrix sizes) at all frequencies. This suggests that we may wish to represent our crack shape/rotation with a non-linear sampling increments i.e. a smaller increment step for longer defects. We also observe that for smaller sub-matrix sizes at higher frequencies the minimum average SSIM score is low. This is best explained by way of considering two small (e.g. 72°) sub-matrices, S_1 & S_2, of the same crack length defect, where one represents a defect with rotation α_{S1} and the other $\alpha_{S1} + 2°$ (where 2° is our angle increment size in our database). S_1 & S_2 are immediate neighbours in our database. Physically the two sub-matrices, S_1 & S_2, represent very similar crack geometries, the only difference being a rotation difference of 2°. We would therefore expect the SSIM score between S_1 & S_2 to be high, which we do not always observe for small sub-matrix sizes at high frequency. The reason for this disparity is twofold. Firstly, the SSIM is sensitive to shift (a translation of spatial coordinates) [16]. Secondly, S_1 & S_2 potentially contain significantly different information content (more pronounced at higher frequency and length, Figure 2b,c) where one of them may potentially encompass an area with greater image complexity and/or spatial information, as shown by the two sub-matrices in Figure 1b.

3.2 Longer range similarity

To explore S-matrix similarity further we analysed the SSIM between a number of S_D within our database. Given that we have 5400 S-matrices, per frequency and sub-matrix size, and are comparing 2, the total number of unique comparisons is 14,577,300, as given by, $P(g,h) = g!/[h!\,(g-h)!]$. Where P is the number of unique combinations, g is the number of elements within our set and h is the number of elements which are being compared. Given the large number of comparisons we resort to sampling a random fraction of the total number of possible comparisons (≈30k). A possible cause of false-positives (which would be the case where two distinctly different cracks are found to have a high SSIM score) in our comparisons will arise from the S-matrix for a given crack with a given rotation angle, α, being identical to a crack of $\alpha \pm 180^{\circ}$ rotation. In order to simplify our analysis we only compare the S-matrices for scatterers of rotations between 0° and 90°, shown in Figure 4. Given the information content shown in Figure 2b,c we would expect the SSIM score to drop as the 'distance' between the two compared S-matrices increases, the physical meaning being that as the 'distance' increases the cracks they represent are increasingly different, in crack length and/or orientation. As indicated in Figure 3, at high frequency and long crack lengths (at all sub-matrix sizes) two neighbouring S-matrices may have significantly different SSIM scores from one another even though they represent physically similar defects.

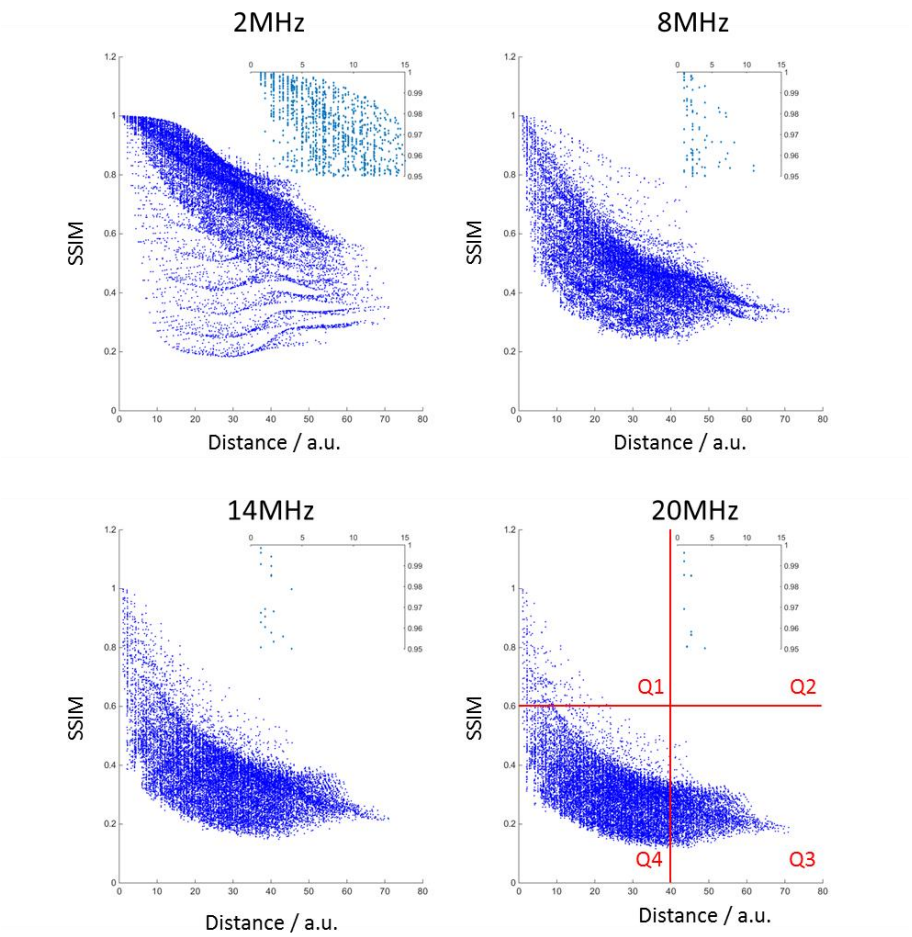

Figure 4 - SSIM and distance, L, for 360° S-matrices. Each frequency shows a zoomed portion of the data for small distances and high SSIM values. Short distances and low SSIM scores represent physically similar defects being significantly different, as implied by their SSIM score.

In Figure 4 it is worth considering what each of the 4 quadrants represent. (Q1) signifies two similar S-matrices which are close within the database, (Q2) signifies two similar S-matrices which are far apart, (Q3) signifies two dissimilar S-matrices which are far apart and (Q4) signifies two dissimilar S-matrices which are close together. Each of these SSIM-distance relations can be explained by the information content (Figure 2b,c) and nearest neighbour (Figure 3) properties of S-matrices.

4 Measurement confidence bounds

We can now address the question: what is the physical meaning of the SSIM score? An SSIM score of 1 would mean that the two S-matrices being compared were identical. As our database is a discrete

representation of possible defect shapes, an SSIM of 1 is unlikely, we thus investigate what non-unity SSIM scores mean. If we were to acquire an experimental S-matrix and compare it to all of the elements of the database, we will be presented with a number of SSIM scores. We could simply make a note of the maximum SSIM score and the corresponding S_D and extract the defect parameters. The problem with this approach is that it gives us no indication of the confidence/error of the result. There will always be a maximum SSIM score even if it is a very low one, the physical meaning being that the defect features corresponding to S_D may be very different from the real defect features.

The confidence in our characterisation result can now be measured by analysing the SSIM distributions shown in Figure 4. As an example, consider a brute force comparison of our experimental S-matrix to all S_D elements of the database and the maximum SSIM score is 0.95. We use this value as it is comparable with the higher SSIM values observed experimentally in [11]. The characterisation error is then defined by the lower and upper bounds of the intersection of the SSIM=0.95 line with the distance data shown in Figure 5a. From this it can be seen that the maximum error (upper bound) in terms of distance is ≈12, which equates to an uncertainty of $0.36\ \mathrm{mm}$, $24°$or some combination thereof. Note from Figure 4 that, the higher the SSIM score, the lower the measurement error will be. Figure 5b shows how the maximum distance error (for SSIM=0.95) varies as a function of sub-matrix size and frequency, from which it can be seen that there is a plateaux of low errors for sub-matrix sizes greater than 90° and frequencies greater than 7MHz. Figure 5b can be seen to act as a guide of the requirements of an array capable of characterising this range of defects to a defined level of accuracy.

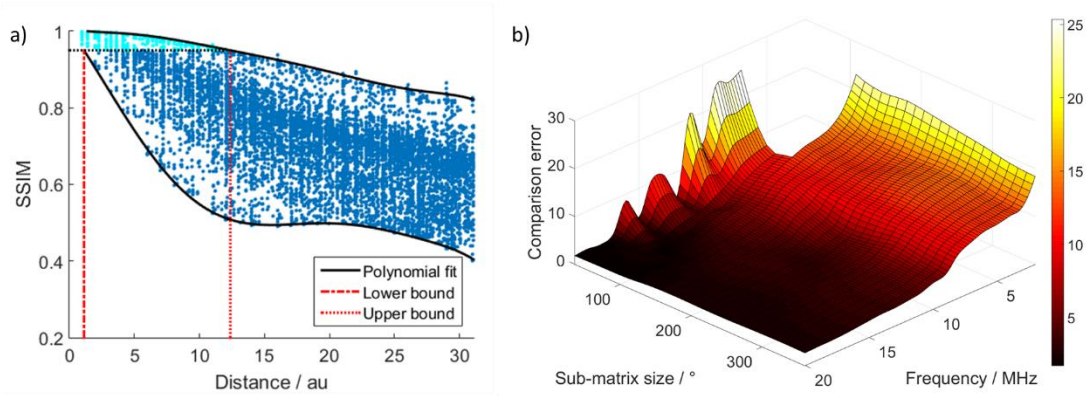

Figure 5 – Confidence measurement. ***a)*** *Data at 5MHz with a sub-matrix size of 360°. A 10th order polynomial line of best fit is plotted and is used to calculate the associated confidence for a given SSIM (in this example the SSIM=0.95 line is shown). The S-matrices above the threshold are shown in cyan, and S-matrices below in blue.* ***b)*** *Upper-bound distance error as a function of frequency and sub-matrix size for an example SSIM score of 0.95. This data can be regarded as a look-up table of maximum possible comparison error. Portions where there is no data (low frequency small sub-matrix size) mean that an SSIM value 0.95 has a comparison error greater than the maximum possible error, in effect, the maximum comparison error is infinite. Interpolation used for* ***b)****.*

5 Influence of scattering noise

The noise included in an experimental measurement may come from many sources and either be incoherent or coherent in nature. Incoherent (or random noise) is less of a concern in NDT as it can be efficiently reduced by averaging the stationary signals [28]. Conversely, coherent noise due to, for example scattering from grain boundaries in metals, is the major limiting factor that determines defect detectability in many real NDT scenarios.

5.1 Noisy S-matrix calculation

In the following we describe a simplified yet realistic noise model based on a Born series to study the effect of scattering noise on a defect's S-matrix. Here we assume that the noise originates from a distribution isotropic point scatterers (these can be thought of as the grains) and solve the multiple scattering process by reducing it to a linear system of equations by means of the self-consistent approach, as in [29]. Consider N isotropic point scatterers centred at $\boldsymbol{r}_1, \boldsymbol{r}_2$, ….etc. In the

neighbourhood of each point scatterer, the scattered field will behave as $E_j G(\boldsymbol{r}, \boldsymbol{r}_j)$, $j = 1,2, \ldots$, where E_j is an unknown complex amplitude and $G(\boldsymbol{r}, \boldsymbol{r}_j)$ is the free space Green's function at a point $\boldsymbol{r}$ in the field of a source at $\boldsymbol{r}_j$. The multiple scattering process between the defect and the point scatterers is presented in the following form

$$u(\boldsymbol{r}) = u^{in}(\boldsymbol{r}) + E_D G(\boldsymbol{r}, \boldsymbol{r}_0) + E_1 G(\boldsymbol{r}, \boldsymbol{r}_1) + E_2 G(\boldsymbol{r}, \boldsymbol{r}_2) + \cdots \quad (3)$$

The scattering properties of the scatterers are characterized by $E_j = s_j u_j^{ex}(\boldsymbol{r}_j)$, where s_j denotes a scattering coefficient. The scattering strength for the defect E_D is determined by the value of the exciting field at $\boldsymbol{r}_0$ together with the amplification factor S_D, that is $E_D = S_D u_0^{ex}(\boldsymbol{r}_0)$. Here, S_D is the response of the defect when it exists alone, *i.e.* the S-matrix. We assume that the scattering coefficients s_j are known. The exciting field u_0^{ex} is the field incident on the defect, in presence of the other scatterers. Evaluation of the exciting field at $\boldsymbol{r}_0$, where the defect is located, yields

$$u_0^{ex}(\boldsymbol{r}_0) = u^{in}(\boldsymbol{r}_0) + \sum_{j=1}^{N} s_j u_j^{ex}(\boldsymbol{r}_j) G(\boldsymbol{r}_0, \boldsymbol{r}_j). \quad (4)$$

Equation 4 contains all multiple scattering interactions between the defect and the point scatterers and is valid for any incident wave $u^{in}(\boldsymbol{r})$ and in any dimension. For simplicity we consider next only the first-order solution of this equation, that is when $u_j^{ex}(\boldsymbol{r}_j) = u^{in}(\boldsymbol{r}_j)$. Replacing the exciting field at the scatterers by the incident field corresponds to the Born approximation [30]. Then, the scattered field at the defect is calculated from

$$u_0^{sc}(\boldsymbol{r}) = S_D u_0^{ex}(\boldsymbol{r}_0) G(\boldsymbol{r}, \boldsymbol{r}_0). \quad (5)$$

This in turn, defines the response of the defect in the presence of 'noise', *i.e.* the 'noisy' S-matrix, S_E, in the far field

$$S_E = U_0 \lim_{\boldsymbol{r} \to \infty} u_0^{sc}(\boldsymbol{r}) / u^{in}(\boldsymbol{r}), \quad (6)$$

where U_0 is a proportionality constant.

In order to calculate S_E, consistent with the definition of S_D (which here assumes 2-D geometries), we consider plane longitudinal waves of angular frequency ω, propagating with wavenumber k, and propagation is modelled using the 2-D free-space Green's function

$$G(\boldsymbol{r}, \boldsymbol{r}_j) = -\frac{i}{4} H_0^{(1)}\left(k|\boldsymbol{r} - \boldsymbol{r}_j|\right), \tag{7}$$

where $H_0^{(1)}$ is a cylindrical Hankel function. The scattering coefficients s_j are determined by imposing appropriate boundary conditions on the surface of each scatterer; they are complex-valued and frequency dependent. In the remainder, we assume that all the scatterers are tiny identical circular inclusions of diameter b, so that

$$s \equiv s_j = i\pi \left(\frac{\kappa_m - \kappa_s}{\mu_m + \kappa_s}\right) \widetilde{\omega}^2, \quad \text{where} \quad \widetilde{\omega} = kb \ll 1. \tag{8}$$

Here, κ and μ denote the bulk and shear moduli, and subscripts 'm' and 's' are related to the host matrix and the scatterers. Scattering is caused by the deviation $\Delta\kappa$ of the bulk modulus of the scatterer κ_s from that of the host matrix κ_m, $\Delta\kappa = \kappa_m - \kappa_s$. In the following, we shall use this quantity to 'tune' the level of the scattering noise. The maximum magnitude of these perturbations was chosen such that the resulting S_E appeared to contain almost no spatially discernible features and were thought to represent noise only. To further support this choice, in the next section we illustrate the effect of noise on imaging of such defects. Note that the size of the scatterers, b, could equally be used as a tuning parameter for fixed $\Delta\kappa$. The 'noise' would therefore increase with the scattering parameter, $\widetilde{\omega} = kb$.

In total, we consider two thousand ($N = 2000$) scatterers (where the scatterers are one micron in size, *i.e.* $b = 1\ \mu\mathrm{m}$) which we randomly place in a circular region of an effective radius $R_{eff} = 25\lambda_M$; here, λ_M denotes the wavelength associated with the smallest frequency (*e.g.* at 2MHz, $\lambda_M \backsimeq 3.1\ \mathrm{mm}$). This implies a constant number density of scatterers, $n_0 = N/(\pi R_{eff}^2)$. The defect is placed at the origin of the region. We perform simulations by solving the wave equations (i.e. 4, 5 and 6) in the far field of the defect, for a single realization of the scatterers' locations. The defect and

the scatterers are illuminated with a plane wave $u^{in} = \exp(ikr\cos\theta_{in})$ at constant incident angle θ_{in},. The same angular sampling is used as for the calculation of S_D.

It is clear that the impact of the distribution of scatterers is such that the waves and spectra associated with a single realization may not approximate a configurationally average wave. However, this is the main attraction of the noise model described here: it provides a simple and realistic representation of scattering noise which may be encountered during inspection of engineering metals – waves propagating in such a noisy medium, would therefore 'see' a single realization of the grains' locations.

5.2 Noisy S-matrix database

To simulate characterising a defect based on its S-matrix in the presence of coherent noise, we create a separate database of noisy S-matrices. This database contains the S-matrices corresponding to 100 unique defects, the lengths and rotations of which are not represented in the original (noise free) database. These simulated noisy S-matrices, form an experimental test dataset and are referred to as S_E. They represent cracks with randomised lengths and rotations between 0.2-2.0 mm and 0-180°, respectively. The S_E database has 100 S-matrices per frequency at 10 sub-matrix sizes (as used in the S_D database) at 10 levels of scattering noise amplitude, including a noise free case.

We then perform a brute force comparison between each S_E (for a given frequency, sub-matrix size and noise level) and the corresponding 5400 S_D elements using the SSIM. The maximum SSIM score indicates which of the S_D elements is most similar to the S_E. To investigate the accuracy of our brute force search of the database we compare the estimated crack parameters for S_E to the true values. We do this for both the crack length and rotation errors, such that

$$\Delta a = \left| a_{S_E} - a_{S_D}^{\max SSIM} \right| \tag{9}$$

$$\Delta \alpha = \left| \alpha_{S_E} - \alpha_{S_D}^{\max SSIM} \right|, \tag{10}$$

where; The a_{S_E} is the experimental defect crack length, $a_{S_D}^{\max SSIM}$ is the database defect crack length for the maximum SSIM score resulting from the brute force search, similarly we calculate the error in rotational angle, α.

5.3 Coherent noise effect on S-matrix features and imaging.

The effect of scattering noise on an S-matrix is illustrated in Figure 6a. Here, an example defect S-matrix is displayed for a range of frequencies and noise levels. As the scattering level increases we see a visual degradation of the noise-free S-matrix features. This manifests itself as increasing localised high/low amplitude 'hot-spots' which appear similar in spatial distribution to white-noise and as well as reduction in edge contrast of the spatial features.

Alongside these results, in Figure 6b we also show representative images of the same defect obtained with using common-source method (CSM) imaging [31]. CSM imaging uses data collected at a set of receiver positions with a common source illuminating the subsurface; it uses all elements to transmit simultaneously, so that plane wave propagates into the structure and an image is created from the scattered waves [31]. The forward model in the previous section was used to simulate the operation of a planar 1-D array with 1500 elements. The output of each element was a five cycle, Gaussian windowed tone-burst with various centre frequencies (*i.e.* 2, 8, 14 and 20 MHz) and a −6 dB bandwidth of 50%. The inspected region is a rectangle of effective sides x_{eff}, z_{eff}. The array extends across the entire length x_{eff} of the region and is located at $(x, z = 0)$. We chose $z_{eff} = 40$ mm, whereas x_{eff} is such that $x_{eff} = \pi R_{eff}^2 / z_{eff}$. Here again, two thousand ($N = 2000$) scatterers are randomly distributed in the rectangle area, with the same number density $n_0 = N/(x_{eff} z_{eff})$ as previously used. The defect is located 20 mm below the surface.

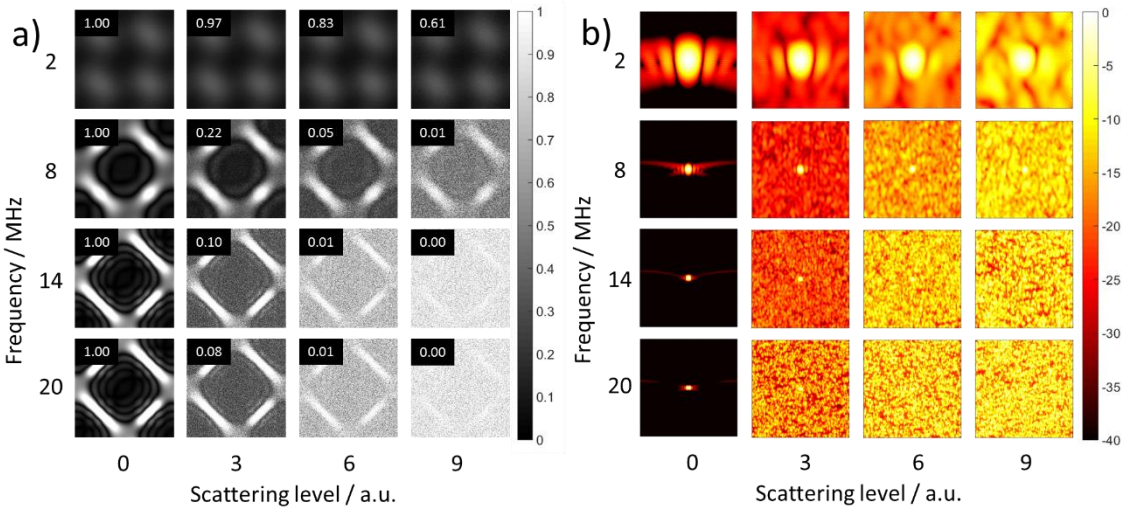

Figure 6 – The effect of simulated scattering noise on S-matrix features and imaging for an example S_E defect (a=0.91mm, α = 0°). ***a)*** *360° S-matrices (angle scale bars omitted for clarity, but are identical to Figure 1b) with SSIM score for comparison to the noise free S-matrix (scattering level= 0).* ***b)*** *CSM images (scale in dB) of 20mm x 20mm area around a centred defect.*

Figure 6a displays, as expected, a significant degradation to the S-matrix with increasing scattering level; the latter was tabbed to increase with increasing Δk in equation 8. The maximum value of Δk was chosen arbitrarily to 200 (the corresponding 'scattering level' is 9 in the figure). The effect of this scattering noise on the CSM images is shown in Figure 6b where the ability to visually identify the defect is reduced at higher frequencies and scattering levels. Interestingly we observe a general trade-off between defect resolution (increasing with frequency, as seen for scattering level = 0) and image noise. Lower frequencies allows the defect location to be identified even at the highest level of scattering noise, but as expected this results in a large point spread function which would mean relatively poor resolving power. At higher frequencies however the defect can be very accurately located, but is only visible above the noise at the lower levels of scattering noise levels.

5.4 Effect of coherent noise on characterisation accuracy

In order to quantify the effect of coherent scattering noise on defect characterisation we perform brute force searches of our noise free database (S_D elements) for our 100 S_E at each frequency and noise level. We choose to only perform this study for a single sub-matrix size of 72° as this is

representative of the angular range one might encounter experimentally [11], [32], [33]. Figure 7 shows the average and maximum errors in estimates of a and α for all 100 S_E, calculated using equations 9 and 10 respectively.

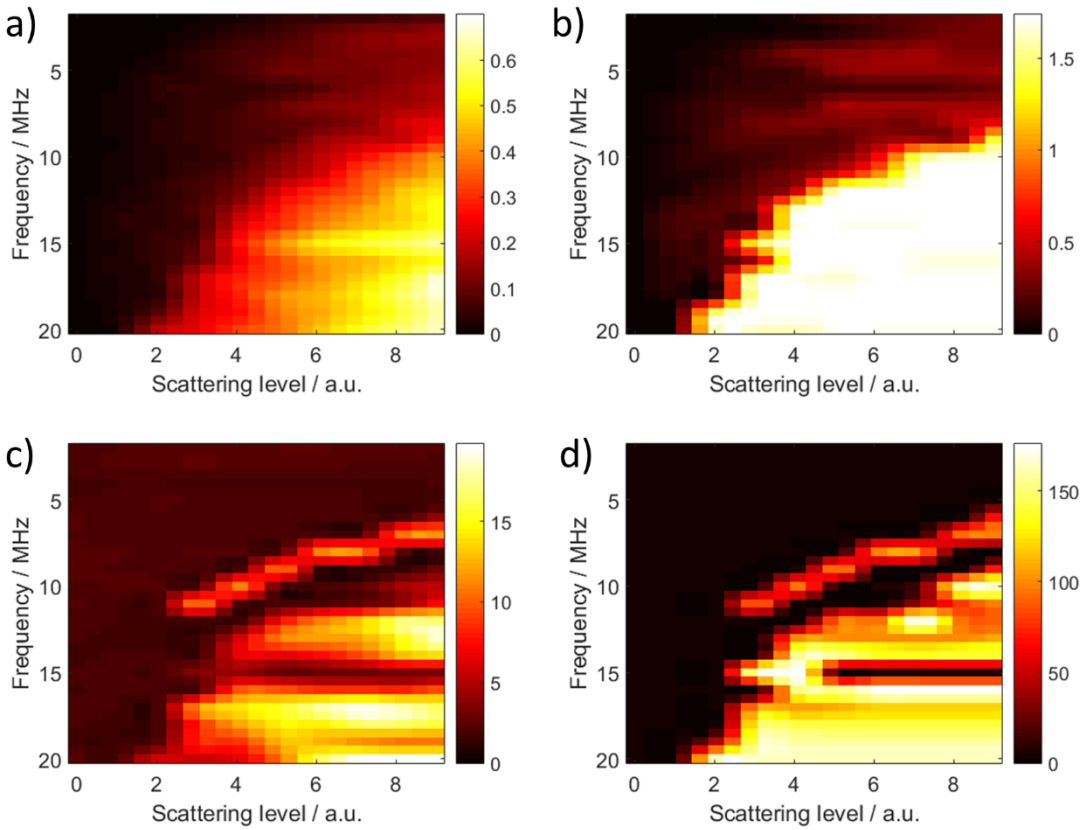

Figure 7 - Effect of scattering noise on characterisation accuracy. ***a)*** *Average crack length error (in mm).* ***b)*** *Maximum crack length error (in mm).* ***c)*** *Average crack rotation error (in deg).* ***d)*** *Maximum crack rotation error (in deg). Interpolation used for all panels.*

Note that the maximum possible error in a or α results when the extracted length is at the extreme values of the database from (i.e. 0.2 mm or 2.0 mm for length and 0° or 178° for angle). For example, if the true length is 0.2 mm and the best match is at 2mm then the maximum error is 1.8 mm. At higher frequencies and scattering levels we observe that many of the maximum errors are at the maximum level. This indicates that at these noise levels/frequencies the defect characterisation *via* a S-matrix database search is ineffective. Conversely, Figure 7 also shows large regions at lower frequencies and scattering levels where the performance of the database search is highly effective.

5.5 Effect of coherent scattering noise on information content

Figure 7 shows that the accuracy of defect characterisation is reduced with increasing scattering level. Here we explore the reason for this based on the information content of S_E with increasing scattering level. Figure 8 shows the variation of image complexity and spatial information of the noisy scattering matrices. Figure 6a shows that increasing noise adds more granular noise to the S-matrix, the effect of which twofold. Firstly, it adds to the image complexity of the S-matrix (top panels in Figure 8) without adding any meaningful data for defect characterisation. In effect the nature of the coherent noise itself increasingly dominates the information as scattering levels or frequency increase. Secondly, the presence of the coherent noise degrades the form of the spatial information seen in Figure 6a, which are the basis for the SSIM comparisons. This reduction in spatial information is clearly shown in the lower panels of Figure 8. The combined effect of the increasing coherent noise is therefore the progressive dominance of the information associated with the noise relative to that associated with the defect.

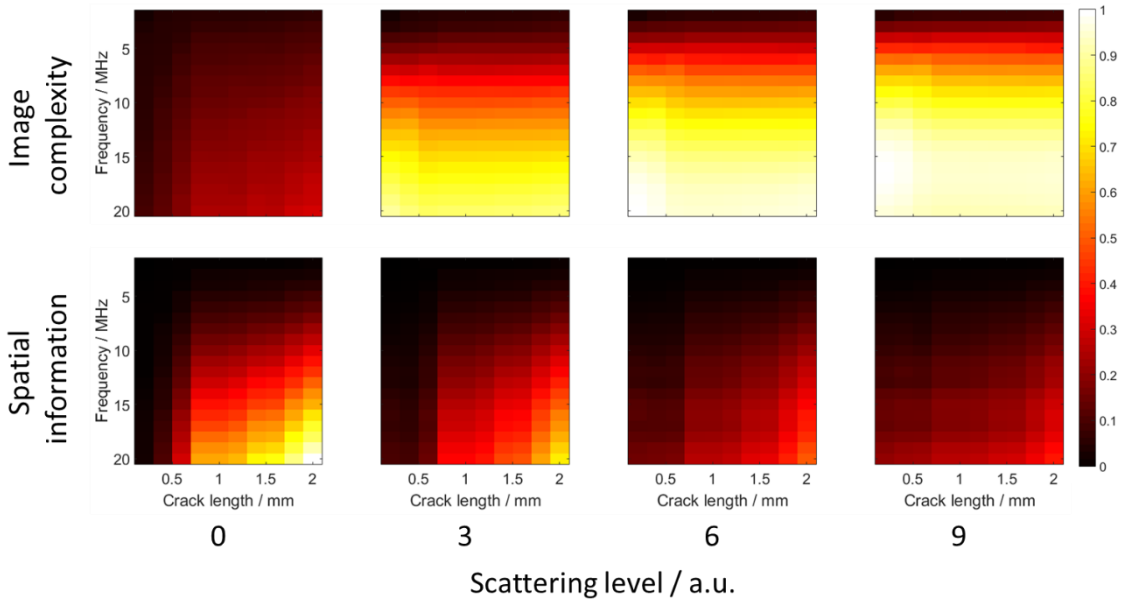

Figure 8 - Effect of scattering noise on S-matrix information content. Image complexity: Normalised file size of S-matrices after JPEG compression. Spatial information: Normalised average number of edges (as calculated using the Sobel operator) of S-matrices. Interpolation used in all panels.

6 Conclusion

S-matrices have been shown to encode information about the geometric parameters of a crack-like defect. Assessment of this S-matrix via similarity and information metrics has shown potential to reveal these features. We have used a large database of pre-calculated S-matrices against which we compare a simulated experimental S-matrix to estimate the geometry of the defect. The nearest neighbour was found using the SSIM metric and in a noise free environment this was shown to work well. By modelling the effects of coherent scattering noise on an S-matrix it has been shown that these estimates of the defect geometry are increasingly inaccurate as the scattering noise increases in magnitude. Coupled with this we have generated a method of estimating the error in our characterisation scheme via an exhaustive search. Analysis of the information content of an S-matrix, has allowed us to understand why the SSIM approach to defect characterisation reduces in accuracy in the presence of coherent scattering noise – in effect the information associated with the coherent noise becomes dominant. An important result of this paper is that the limitations inherent in characterisation via a defects' S-matrix have been quantified. For example, the paper highlights regions of the parameter space where, if the array is designed correctly, good characterisation results are possible even in the presence of coherent noise. Future work will explore the extension of this database search approach to a wider range of defects.

7 Acknowledgements

This work was funded through the UK Research Centre in NDE (RCNDE) and by the UK Engineering and Physical Sciences Research Council (EPSRC).

8 References

[1] C. Holmes, B. W. Drinkwater, and P. D. Wilcox, “Post-processing of the full matrix of ultrasonic transmit–receive array data for non-destructive evaluation,” *NDT E Int.*, vol. 38, no. 8, pp. 701–711, Dec. 2005.

[2] A. N. Sinclair, J. Fortin, B. Shakibi, F. Honarvar, M. Jastrzebski, and M. D. C. Moles, “Enhancement of ultrasonic images for sizing of defects by time-of-flight diffraction,” *NDT E Int.*, vol. 43, no. 3, pp. 258–264, Apr. 2010.

[3] M. Spies and H. Rieder, “Synthetic aperture focusing of ultrasonic inspection data to enhance the probability of detection of defects in strongly attenuating materials,” *NDT E Int.*, vol. 43, no. 5, pp. 425–431, Jul. 2010.

[4] N. Portzgen, D. Gisolf, and G. Blacquiere, “Inverse wave field extrapolation: a different NDI approach to imaging defects,” *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 54, no. 1, pp. 118–127, Jan. 2007.

[5] A. J. Hunter, B. W. Drinkwater, and P. D. Wilcox, “The wavenumber algorithm for full-matrix imaging using an ultrasonic array,” *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 55, no. 11, pp. 2450–2462, Nov. 2008.

[6] J. Zhang, B. W. Drinkwater, and P. D. Wilcox, “The use of ultrasonic arrays to characterize crack-like defects,” *J. Nondestruct. Eval.*, vol. 29, no. 4, pp. 222–232, Sep. 2010.

[7] K. M. M. Tant, A. J. Mulholland, and A. Gachagan, “A model-based approach to crack sizing with ultrasonic arrays,” *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 62, no. 5, pp. 915–926, May 2015.

[8] L. Satyanarayan, A. Kumar, T. Jayakumar, C. V. Krishnamurthy, K. Balasubramaniam, and B. Raj, “Sizing Cracks in Power Plant Components Using Array Based Ultrasonic Techniques,” *J. Nondestruct. Eval.*, vol. 28, no. 3–4, pp. 111–124, Dec. 2009.

[9] R. Demirli and J. Saniie, “Model-based estimation of ultrasonic echoes. Part I: Analysis and algorithms,” *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 48, no. 3, pp. 787–802, May 2001.

[10] R. Demirli and J. Saniie, “Model-based estimation of ultrasonic echoes. Part II:

Nondestructive evaluation applications," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 48, no. 3, pp. 803–811, May 2001.

[11] L. Bai, A. Velichko, and B. Drinkwater, "Ultrasonic characterization of crack-like defects using scattering matrix similarity metrics," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 62, no. 3, pp. 545–559, 2015.

[12] J. D. Achenbach, "Quantitative nondestructive evaluation," *International Journal of Solids and Structures*, vol. 37, no. 1–2. pp. 13–27, 2000.

[13] P. A. Martin, *Multiple Scateering - Interaction of Time-Harmonic Waves with N Obstacles*, 1st ed. Cambridge: Cambridge University Press, 2006.

[14] A. Velichko and P. D. Wilcox, "Reversible back-propagation imaging algorithm for postprocessing of ultrasonic array data," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 56, no. 11, pp. 2492–2503, Nov. 2009.

[15] L. Bai, A. Velichko, and B. W. Drinkwater, "Characterization of defects using ultrasonic arrays: a dynamic classifier approach.," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 62, no. 12, pp. 2146–60, Dec. 2015.

[16] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.

[17] A. Velichko and P. D. Wilcox, "A generalized approach for efficient finite element modeling of elastodynamic scattering in two and three dimensions.," *J. Acoust. Soc. Am.*, vol. 128, no. 3, pp. 1004–1014, 2010.

[18] M. T. Hutchings, P. J. Withers, T. M. Holden, and T. Lorentzen, *Introduction to the Characterization of Residual Stress by Neutron Diffraction*. Boca Raton, Florida: CRC Press, 2005.

[19] J. Rigau, M. Feixas, and M. Sbert, "An information-theoretic framework for image complexity," in *Computational Aesthetics'05 Proceedings of the First Eurographics conference on Computational Aesthetics in Graphics, Visualization and Imaging*, 2005, pp. 177–184.

[20] A. N. Kolmogorov, "On tables of random numbers," *Theor. Comput. Sci.*, vol. 207, no. 2, pp. 387–395, Nov. 1998.

[21] R. Cilibrasi and M. B. Vitanyi, “Clustering by compression,” *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1523–1545, 2005.

[22] G. K. Wallace, “The JPEG still picture compression standard,” *Commun. ACM*, vol. 34, no. 4, pp. 30–44, Apr. 1991.

[23] Y. Honghai and S. Winkler, “Image complexity and spatial information,” *Qual. Multimed. Exp. (QoMEX), Fifth Int. Work. on. IEEE*, 2013.

[24] A. Wenzel, E. Gotfredsen, E. Borg, and H. G. Gröndahl, “Impact of lossy image compression on accuracy of caries detection in digital images taken with a storage phosphor system.,” *Oral Surg. Oral Med. Oral Pathol. Oral Radiol. Endod.*, vol. 81, no. 3, pp. 351–355, 1996.

[25] A. Janhom, P. F. Van der Stelt, and F. C. Van Ginkel, “Interaction between noise and file compression and its effect on the recognition of caries in digital imaging,” *Dentomaxillofacial Radiol.*, vol. 29, no. 1, pp. 20–27, 2000.

[26] ANSI T1.801.03-2003, “Digital Transport of One-Way Video Signals – Parameters for Objective Performance Assessment Language.” American National Standard for Telecommunications.

[27] B. V Dasarathy, *Nearest Neighbor: Pattern Classification Techniques*. Michigan, USA: IEEE Computer Society Press, 1990.

[28] S. V. Vaseghi, *Advanced Digital Signal Processing and Noise Reduction*, 4th ed. Chichester, England: Wiley, 2008.

[29] L. L. Foldy, “The Multiple Scattering of Waves. I. General Theory of Isotropic Scattering by Randomly Distributed Scatterers,” *Phys. Rev.*, vol. 67, no. 3–4, pp. 107–119, Feb. 1945.

[30] M. Fiddy and R. S. Ritter, *Introduction to Imaging from Scattered Fields*, 1st ed. Boca Raton: CRC Press, 2014.

[31] R. Y. Chiao and L. J. Thomas, “Analytic evaluation of sampled aperture ultrasonic imaging techniques for NDE,” *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 41, no. 4, pp. 484–493, 1994.

[32] J. Zhang, B. W. Drinkwater, and P. D. Wilcox, “Longitudinal wave scattering from rough crack-like defects,” *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 58, no. 10, pp. 2171–2180, 2011.

[33] J. Zhang, B. W. Drinkwater, and P. D. Wilcox, “Defect characterization using an ultrasonic array to measure the scattering coefficient matrix,” *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 55, no. 10, pp. 2254–2265, 2008.